



CONSOLIDER LES DONNEES POUR LA GESTION ET LA DIFFUSION DES COLLECTIONS D'UN MUSEE DE FRANCE



Mise en ligne : 4 novembre 2016

Introduction : que recouvre la consolidation des données ?

1 - Objectifs

2 - Opportunité de la consolidation des données

3 - Précautions nécessaires

4 - Processus de consolidation

4.1- Expertise préalable

4.2 - Identification du périmètre des données à consolider

4.2.1 - Les index

4.2.2 - Les notices

4.2.2.1 - La structure des notices

4.2.2.2 - Le contenu des notices

4.2.3 - Les images

4.2.4 - L'administration de la base

4.2.4.1 - Les profils

4.2.4.2 - La gestion des droits

4.3 - Définition des priorités d'intervention

4.3.1 - Evaluation de l'urgence et de l'importance des problèmes à résoudre

4.4 - Organisation des interventions

5 - Clôture de la consolidation des données

Introduction : que recouvre la consolidation des données ?

Exploiter des données plus solides, plus fiables, plus pertinentes, voici un objectif partagé par beaucoup d'administrateurs de bases de données de musées. L'enjeu est d'avoir à disposition un système informatisé qui reflète la réalité administrative, scientifique et documentaire des collections conservées.

Or, même si on dispose d'une charte de saisie, la base de données peut, au fil du temps et des saisies multiples voire stratifiées, perdre en cohérence, en exactitude et en homogénéité. Ces évolutions ont souvent des conséquences négatives sur la qualité de la recherche et pénalisent le système d'information.

Le fait de consolider les données va tenter de remédier à cette perte de qualité et à rendre la recherche et l'exploitation des données plus efficaces. Concrètement, cela recouvre les interventions raisonnées, opérées sur l'ensemble de la base, pour rectifier la graphie, la syntaxe, la structure ou le contenu (fond) de certaines informations.

Ces interventions seront programmées en évaluant :

- l'importance et l'urgence des problèmes,
- la possibilité de les résoudre.

Il peut s'agir de corrections, suppressions, actualisations, ajouts ou conversions d'informations.

Elles peuvent être faites à différents niveaux de la base de données : index (termes doublons, fautes d'orthographe), notices (harmonisation de la structure, de la forme et du contenu), administration de la base (adaptation du paramétrage et des droits d'accès).

L'opportunité et la faisabilité de cette consolidation est évaluée à partir d'un état des lieux des données et de l'utilisation de la base au sein du musée. A cette fin, utiliser le Vade-mecum de préparation des données.

1 - Objectifs

Améliorer la qualité des informations (fond et forme) permettra, de façon plus ou moins immédiate, en fonction des nécessités, de **fiabiliser la recherche** et de la rendre plus pertinente.

Des données cohérentes sont interrogeables sans générer de bruit (toute réponse non pertinente à une recherche documentaire), ni de silence (des documents pertinents et existant dans la base de données ne sont pas sélectionnés à la suite de l'interrogation).

L'un des gains attendus de ce travail est de **rendre l'information réellement exploitable** sur toute la chaîne documentaire. La consolidation d'un ensemble homogène de rubriques constituant le socle commun minimal d'une majorité de notices permettra d'en bonifier l'exploitation en interne jusqu'à la mise en ligne.

2 - Opportunité de la consolidation des données

Quel que soit le contexte qui justifie une consolidation des données, il faut avoir à l'esprit que ces interventions globales nécessitent bien souvent d'**immobiliser la base** durant un certain laps de temps. Ceci est nécessaire afin de ne pas créer d'interférence entre les données. Cette durée doit être évaluée à l'avance afin de ne pas pénaliser le travail des personnes intervenant ou consultant la base. Ces dernières devront être prévenues de l'interruption mais également de la remise en service du système.

A quels moments peut-on consolider les données ?

- **avant une reprise des données** : une consolidation, intégrée à la préparation des données transmises au prestataire, rendra la reprise des informations moins compliquée et donc moins coûteuse.

Il est fréquent que des utilisateurs aient, au fil des années, contourné les règles de saisie orthodoxe du logiciel . Ces décalages ne seront pas forcément anticipés par le prestataire dont le pré-requis est l'utilisation normale du logiciel. Le redressement de ces décalages nécessitera une analyse plus poussée du prestataire afin de proposer un scénario de reprise qui sera forcément plus onéreux. La consolidation des données par le musée, avant la reprise, permettra de limiter les interventions du prestataire, et donc les coûts de prestation.

L'objectif de cette consolidation préalable est d'assurer le meilleur fonctionnement du nouveau système lors de son déploiement et de permettre une meilleure adaptation de l'équipe à ce changement important.

- **après un import ou une reprise de données** : des interventions globales peuvent se révéler nécessaires pour régler d'éventuels problèmes issus de la reprise. La préparation des données par le musée et l'analyse du prestataire devraient éviter au maximum les interventions a posteriori ; ces dernières peuvent malgré tout s'imposer pour bénéficier pleinement du nouveau système. (voir reprise des données),

- **à échéance régulière** : une consolidation périodique des données fait partie du bon entretien de la base. Cela relève de l'administration de la base de données,

- **en vue de l'édition des données portées informatiquement au registre d'inventaire** : cette procédure réglementaire impose de fiabiliser préalablement les notices d'un point de vue administratif, juridique et scientifique. Les notices réglementaires sont rendues non modifiables avant leur édition en fin d'année afin de rendre strictement compte de l'état des connaissances sur une œuvre à la date d'inscription sur le registre d'inventaire. Il convient donc d'avoir rigoureusement vérifié que les données réglementaires sont correctes tant au niveau du fond qu'au niveau de la forme avant leur validation et verrouillage,

- **en vue d'une publication / diffusion des données** : les notices vont être, à travers une mise en ligne, exposées largement au public et donc, dans une large mesure, refléter l'image du musée. Le public est en droit de considérer les données délivrées par des institutions culturelles comme des données dites « de confiance ». Il relève donc de la responsabilité des musées de France, en adéquation avec les missions décrites par le Code du patrimoine, de diffuser des données qualifiées et vérifiées. Pour le garantir, il s'avère souvent nécessaire d'améliorer la qualité et la fiabilité des informations avant leur publication. Ceci est vrai quel que soit le support de diffusion, mais d'autant plus impératif lors de la mise en ligne sur un site fédérant des collections (base régionale notamment) dont la présentation, la structure et l'ambition de contenu doivent être harmonisés entre partenaires selon la charte éditoriale retenue.

3 - Précautions nécessaires

La consolidation des données doit s'appuyer sur la **charte de saisie** du musée et les outils de référence traditionnellement utilisés ou plus récemment apparus dans le domaine scientifique dont relèvent les collections du musée (dictionnaire technique, catalogue raisonné, etc.). Il est fréquent que la mise en place d'une charte de saisie soit suivie d'une phase de consolidation des données. En effet, certaines consignes peuvent directement traiter de la correction, progressive et répartie en équipe, de données fautives, issues par exemple d'une saisie ancienne. Et, de même, il peut s'avérer nécessaire de réviser la charte au cours ou à l'issue de la consolidation.

Un principe important de mise en œuvre de cette consolidation est de l'entreprendre selon un **programme bien défini et de ne pas s'arrêter en plein milieu**. Le risque serait de créer des incohérences supplémentaires et donc d'être contreproductif par rapport à l'objectif visé.

Après l'identification des priorités, selon l'organisation du musée, la consolidation des données sera :

- pilotée, voire assurée, par l'**administrateur**,
- répartie selon les forces de l'**équipe**, et le périmètre d'intervention de chacun,
- sous-traitée : un **renfort extérieur** peut être sollicité mais l'analyse des corrections nécessaires doit être faite ou validée par le musée. Le **personnel scientifique** devra être étroitement impliqué dans la correction et la validation de données ayant trait au contenu des notices.

Quoi qu'il en soit, toute consolidation nécessite une excellente **connaissance de la base et des pratiques de saisie** ainsi que la maîtrise des **fonctionnalités de recherche avancée et de modification par lots** du logiciel.

L'**assistance technique** de l'éditeur du logiciel sera utilement sollicitée dans le cadre du contrat de maintenance.

Des besoins de **formation** peuvent également être identifiés à cette occasion.

4 - Processus de consolidation

Voici les étapes du processus classique d'une consolidation des données :

- Expertise préalable,
- Identification des problèmes et du périmètre des données concernées,
- Pondération des problèmes,

- Définition du degré de priorité de traitement de chaque problème,
- Définition du programme de résolution des problèmes,
- Déploiement du programme de résolution des problèmes,
- Pointage et suivi des problèmes résolus,
- Pérennisation des bonnes pratiques,
- Bilan d'activité.

Nous détaillons ci-dessous les étapes clés.

4.1 - Expertise préalable

Il est fréquent que l'administrateur ou l'équipe aient parfaitement conscience de ce qui peut poser problème dans la base de données. Néanmoins, une expertise préalable des données s'impose afin d'avoir **une vue d'ensemble de l'état de la base**.

Selon le nombre de notices, on procédera par relecture ou sondages afin de déterminer :

- ce qui est effectivement renseigné,
- l'homogénéité et la qualité des données,
- la valeur informationnelle.

Ce diagnostic qualitatif permettra au musée d'identifier les points faibles et les éléments cruciaux auxquels veiller.

En cas d'export Joconde, il est possible de s'appuyer sur l'analyse de la responsable Joconde qui aura éventuellement mis en lumière des soucis de cohérence internes à la base.

Les problèmes identifiés lors de l'expertise préalable de la base peuvent être de diverses natures :

- orthographe,
- forme,
- syntaxe,
- lexique,
- structure de la notice,
- obsolescence,
- données « pauvres » (notice trop succincte),
- données « fausses » (notice non fiable scientifiquement, administrativement),
- mauvaise couverture photographique de la base (qualité de la prise de vue ou de la numérisation, incomplétude, gestion du droit d'auteur, liens texte-image)
- incohérences ou besoin de régularisation du paramétrage de la base.

4.2 - Identification du périmètre des données à consolider

En fonction du type de problème, l'expertise et les interventions se focaliseront sur certains éléments de la base :

- Les index,
- La structure des notices,
- Le contenu des notices,
- Les images et les liens image-notice,
- L'administration de la base.

4.2.1 - Les index

Les index et les tables répertorient les listes des descripteurs utilisés dans les champs de vocabulaire contrôlé de la base. Ces listes de termes, visualisées dans l'ordre alphabétique ou selon leur organisation hiérarchique, constituent le **point d'entrée privilégié pour expertiser la nature et l'usage des vocabulaires**. En effet, il serait vraiment aberrant d'analyser la qualité des vocabulaires en passant chaque notice en revue alors que les termes sont justement regroupés et centralisés dans les tables, accessibles via l'outil.

Cette expertise passera par le **repérage des termes doublons, fautifs, orphelins ou mal placés dans l'arborescence**. Chaque terme de ces listes sera utilement mis en regard avec le nombre d'occurrences (le nombre de fois où le terme est utilisé dans la base). Ce rapport permet d'évaluer l'impact du problème et le bénéfice attendu de sa résolution.

Selon la gestion des vocabulaires exposée dans la charte de saisie et selon les priorités définies, notamment en fonction du nombre d'occurrences concernées, il sera possible de **corriger ces erreurs ou ces incohérences directement dans les index ou les thesaurus**. Grâce au caractère relationnel des fichiers, corriger un terme via un index a pour conséquence de corriger automatiquement toutes les occurrences de ce terme. Ainsi, c'est également le contenu de la notice qui est amélioré.

Ce travail sur les vocabulaires peut également avoir pour suite l'**insertion de notes d'usage** pour certains termes rares, complexes, techniques ou vernaculaires.

4.2.2 - Les notices

Différents types de notices peuvent nécessiter une consolidation (correction, actualisation) des données qu'elles contiennent.

Quel que soit le cas de figure, il est important de **mémoriser le moyen qui vous permet d'isoler ces notices au cœur de la base** et de les retrouver au moment choisi pour les corriger. Selon les cas, il peut s'agir de l'enregistrement des critères de recherche, de l'ajout d'un marqueur sur ces notices ou encore de l'introduction d'une valeur discriminante, toujours identique, dans un champ destiné à la gestion de la base.

- **Notices anciennes** : l'ancienneté d'une notice dans la base n'est pas forcément un signe de mauvaise qualité. Néanmoins, dans certains musées, la saisie peut dater de l'acquisition de l'outil et ses principes avoir grandement évolué. Cette saisie ancienne peut nécessiter de sérieuses mises à jour. C'est la date d'enregistrement de la notice qui sera le critère permettant de les retrouver afin d'en faire l'expertise et de les corriger (substitution ou ajout d'informations, voire suppression de la notice).

Lors de la reprise de données d'un système informatisé vers un nouveau, on peut avoir à gérer un certain « passif » qui aura été idéalement identifié lors de l'analyse préalable, comptabilisé et pourvu, lors de la migration d'un critère distinctif. Pour retrouver et traiter ces notices dans le nouvel outil, on utilisera donc ce critère.

- **Notices issues d'une saisie mal contrôlée** : il peut s'agir d'erreurs ou d'incohérences de saisie ayant eu cours avant la mise en place de la charte ou issues d'une saisie déléguée, mal encadrée ou non révisée. Pour les isoler, il faut les sélectionner par le dénominateur commun pertinent (nom du rédacteur, thème ou date de saisie, etc.)

- **Notices non validées** : pour des raisons majoritairement liées à l'organisation de l'institution, des notices peuvent présenter une saisie soignée et aboutie mais ne pas avoir été validées par un relecteur final. La non validation de ces notices qui peuvent être fort nombreuses dans certains systèmes, peut constituer un vrai blocage dans l'utilisation partagée de l'outil, la diffusion des informations et la motivation de l'équipe. Cela laisse planer un doute sur la qualité et l'utilité du système d'information à grande échelle. L'ambiguïté réside dans le fait que des notices de bonne qualité peuvent ne pas avoir été davantage validées que des mauvaises. Une réflexion en interne s'impose pour définir des critères de qualité réalistes et mettre au point un circuit de validation efficace, éventuellement allégé. Une fois un dispositif décidé, techniquement, il faut sélectionner ces notices non validées par l'absence de signe de validation (statut de la notice, absence de date ou mention du nom de valideur).

- **Notices que le musée souhaite publier** : les sélectionner selon le thème, la publication choisie et un éventuel critère de diffusion public (conformité au droit d'auteur des images illustrant les notices, notamment).

4.2.2.1 - La structure des notices

Si la structure des notices doit être révisée, les interventions impacteront l'**ordre et l'organisation des rubriques qui composent la notice**. Cela reposera sur une analyse globale de la saisie opérée sur la base, compte-tenu que ces rubriques peuvent être rédigées en texte libre ou à l'aide d'un vocabulaire contrôlé.

Des vérifications seront faites à partir de la charte de saisie. Certaines rubriques n'ont peut-être plus leur raison d'être en fonction des profils de saisie préconisés.

On évaluera l'**homogénéité de la saisie** : certaines rubriques peuvent être systématiquement renseignées ; à l'inverse d'autres peuvent l'être rarement, ou de façon incomplète. On peut alors estimer nécessaire de passer les informations d'une rubrique obsolète vers une rubrique privilégiée ; puis de rendre impossible l'utilisation de cette rubrique obsolète.

Une attention particulière pourra être portée sur les **rubriques rédigées en texte libre** : les laisse-t-on telles quelles ou peut-on les scinder ? Des corrections peuvent être apportées au niveau du **format** de certaines informations, comme les dates, par exemple.

4.2.2.2 - Le contenu des notices

Il s'agit de vérifier la **valeur informationnelle des notices**. Elle consiste en la fiabilité et la qualité des informations administratives, historiques, documentaires, scientifiques ou relatives à la gestion des biens décrits.

Si le contrôle de ces informations ne peut être qu'humain et qualifié, les interventions qui en découlent peuvent être de plusieurs ordres :

- **suppression** des valeurs aberrantes (données obsolètes, inutilisées, doublons, erreurs répétées de saisie),
- **correction** de lacunes d'information (unité de mesure...),
- recherches documentaires pour fiabiliser le fond et **ajout d'informations**,
- mise en place d'un **circuit de validation** pour notices non validées administrativement ou scientifiquement,
- **suppression** ou **changement de statut** de la notice.

En règle générale, il faut faire **attention aux rubriques de commentaires** (notes) ou de précisions et limiter leur usage. En effet, elles peuvent contenir abusivement des informations qui pourraient être plus efficacement retrouvées dans des rubriques composées de termes contrôlés. La facilité avec laquelle on retrouve l'information doit être le critère de décision numéro un pour la saisie et sa consolidation.

4.2.3 - Les images

La question des images est centrale pour assurer la qualité de la base de données.

Le nombre des images sur les collections a énormément augmenté ; cela résulte notamment des prises de vues parfois systématiques lors des campagnes de récolement. Il s'agit de mettre au point une stratégie adaptée au musée permettant de trouver un point d'équilibre entre photographies de travail, photographies commercialisables ou réutilisables.

Des précautions techniques pour fiabiliser l'association notice / image sont également à prendre en compte et entrent dans le périmètre de la consolidation des données.

- expertise des **nommages et répertoires d'images**. Attention, modifier l'arborescence des répertoires de stockage ou les noms de fichiers numériques peut avoir comme conséquence désastreuse de casser les liens qui associent les notices d'œuvres à leur(s) illustration(s). La prudence est donc de mise.
- expertise de la **qualité de prise de vue ou de numérisation** : identifier les lots d'images posant problème (selon la date de prise de vue, la thématique ou l'auteur des clichés).
- expertise de l'**association entre notices et images** : identification et comptage des liens invalides, comptage des notices non illustrées et comparaison avec les fonds d'images existants qui pourraient y être associés sans prise de vue ou numérisation supplémentaires.

Selon les priorités définies, de **nouvelles prises de vue ou campagnes de numérisation** pourront être programmées selon une liste et des consignes précises (définition, résolution, etc.). En fonction des besoins, les **liens images seront restaurés** pour assurer une bonne couverture photographique de la base.

4.2.4 - L'administration de la base

Consolider la base, c'est aussi vérifier que son utilisation est bien régie. Cette vigilance est une des composantes du rôle de l'administrateur de la base (lien). Elle s'opère à plusieurs niveaux.

4.2.4.1 - Les profils

Ces masques de saisie personnalisés sont paramétrés par l'utilisateur ou l'administrateur (selon la gestion des droits définie en amont). Ils regroupent et permettent de ne visualiser à l'écran que les rubriques nécessaires à la saisie d'un groupe d'objets appelant le même traitement.

Il est nécessaire de les identifier et d'**en assurer la bonne gestion, particulièrement si les utilisateurs y ont fréquemment recours et tendance à les multiplier** :

- **Lister** les profils par catégorie : recherche, saisie, édition, impression, publication des données.

- **Conserver** les profils utiles (référencés dans la charte de saisie) et supprimer les doublons ou les profils obsolètes,
- Les **répertoire en détail permettra de les recréer** en cas de problème ou de changement d'outil.

Il ne faut pas oublier que les profils peuvent être déclinés pour la saisie, la recherche, l'édition, l'impression ou encore la publication des données. La cohérence de ces paramètres devra être vérifiée.

4.2.4.2 - La gestion des droits

Le paramétrage des droits en lecture et écriture des différents utilisateurs de la base de données doit être mémorisé et actualisé.

Si cela n'a pas encore été fait, **lister** :

- les groupes d'utilisateurs,
- leurs droits en lecture et écriture sur telle ou telle information (champ ou table).

Les **répertoire en détail permettra de les recréer** en cas de problème ou de changement d'outil.

Les **actualiser** en fonction de l'équipe : nouveaux membres, besoins et périmètres d'intervention.

Selon les constats faits lors de l'expertise préalable à la consolidation, une **révision voire restriction** de certains droits peuvent s'avérer nécessaire afin de limiter le périmètre d'intervention sur la base de certains agents. Si ces changements sont importants, il peut être utile de les faire valider hiérarchiquement et de les communiquer en toute transparence. Un accompagnement personnalisé et/ou une formation complémentaire pourront être proposés éventuellement aux agents concernés.

4.3 - Définition des priorités d'intervention

Une fois les problèmes identifiés en prenant en compte la volumétrie des données concernées, il est nécessaire d'établir un ordre des interventions selon leur priorité respective et d'en assurer le suivi.

Il s'agit de déterminer selon le type de chaque problème

- l'entité documentaire concernée et sa volumétrie,
- le degré d'importance du problème,
- le degré d'urgence de la résolution du problème,
- le moyen de corriger (intervention sérielle automatique, recherche complémentaire),
- la facilité de la correction.

Référez-vous au tableau d'analyse des priorités.

4.3.1 - L'urgence et l'importance des problèmes à résoudre

L'**importance** d'une activité se détermine en fonction de l'impact de la non réalisation de cette activité ("quel est le risque à ne pas faire ?"), de l'exigence hiérarchique voire politique, du caractère obligatoire ou réglementaire de l'information ou du dossier à traiter.

L'**urgence** est déterminée par le délai ou l'échéance qui s'impose à la réalisation de la tâche. Ceci est accru si cette réalisation est le pré-requis à la mise en place d'autres activités (mise en ligne, etc.).

Une tâche prioritaire combine un haut degré d'importance et un haut degré d'urgence.

Prioriser permet donc de se donner les moyens d'accomplir l'essentiel, de synchroniser et de coordonner des activités plus ou moins complexes, plus ou moins intrinsèquement liées à l'administration de la base de données.

Il est possible d'**évaluer les problèmes et de définir les priorités selon la matrice d'Eisenhower.**

Eisenhower aurait un jour déclaré : « *Ce qui est important est rarement urgent et ce qui est urgent rarement important* ».

Quatre niveaux de priorités peuvent être définis en fonction du degré de l'importance et de l'urgence du problème à résoudre ou de l'activité à effectuer :

- A : important et urgent. A corriger immédiatement par l'administrateur = priorité
- B : important et peu urgent. A programmer ou déléguer = potentiel futur problème

- C : urgent et peu important = à corriger rapidement
- D : peu urgent et peu important = à différer ou à abandonner

Quelques exemples dans le domaine de la qualité des données patrimoniales :

A : important et urgent

Le mode d'acquisition est erroné sur un lot de notices qui va être porté sur le registre d'inventaire dans quelques semaines.

Importance : caractère réglementaire de l'information erronée et du document final ; volumétrie des documents

Urgence : délai réglementaire de validation et d'impression du document final

B : important et peu urgent

Les particules des noms d'auteurs ont été mal rejetées dans le lexique. Il faudra régler le problème pour que les notices de ces artistes « sortent » sur la base régionale.

Importance : fiabilité de l'interrogation de la base régionale ; risque de silence et de manque de visibilité des ressources du musée

Urgence : pas de délai associé, néanmoins, il est sans aucun doute plus simple d'opérer cette correction avant l'export des données sur la base régionale

C : urgent et peu important

L'unité de mesure n'est pas précisée dans les dimensions de trois objets dont l'export Joconde est envisagé.

Importance : l'information manquante relève de la précision ; faible volumétrie de documents

Urgence : une date d'export a dû être programmée

D : peu urgent et peu important

Les sigles de certaines institutions partenaires du musée ne sont pas développés. Une note d'usage dans la liste d'autorité et une équivalence pourront être entrées lors du traitement de l'index des personnes morales.

Importance : l'information manquante relève de la précision ; simplicité de résolution transversale à travers les index

Urgence : pas de délai de résolution.

4.4 - Organisation des interventions

Les priorités, qu'elles soient identifiées en équipe ou par l'administrateur de la base de données, peuvent être ensuite reprises dans une feuille de route à faire valider par sa hiérarchie, ainsi qu'un planning prévisionnel.

La liste des interventions à mener, composante majeure de la feuille de route, listera, pour chaque problème à résoudre :

- le niveau de priorité,
- l'intervention à mener (moyen retenu pour solutionner le problème),
- la date de prise en charge du problème,
- acteur prenant en charge le problème,
- date prévisionnelle de prise en charge du problème ; périodicité,
- date effective de prise en charge du problème,
- date de résolution du problème,
- commentaire.

Cette liste permettra d'organiser le travail de consolidation, en fonction des priorités, en équipe et de programmer un calendrier prenant en compte le séquençage des activités et leurs différentes dates butoir.

5 - Clôture de la consolidation des données

Une fois les interventions pour consolider la base, opérées, il convient d'en faire le bilan et d'en garder la trace.

L'actualisation de la charte de saisie permettra de pérenniser les bonnes pratiques, qu'il s'agisse de nouvelles méthodes et d'interventions déléguées au reste de l'équipe. N'oubliez pas qu'une base de données documentaire est un outil vivant, amené à évoluer continuellement et sans cesse perfectible.